CLCS : Contrastive Learning between Compositions and Structures for practical Li-ion battery electrodes design

Jaewan Lee
LG AI Research
Seoul, Korea
jaewan.lee@lgresearch.ai

Changyoung Park
LG AI Research
Seoul, Korea
changyoung.park@lgresearch.ai

Hongjun Yang
LG AI Research
Seoul, Korea
hongjun.yang@lgresearch.ai

Sehui Han
LG AI Research
Seoul, Korea
hansse.han@lgresearch.ai

Woohyung Lim
LG AI Research
Seoul, Korea
w.lim@lgresearch.ai

Abstract

Prediction of average voltage of a cathode material, which is related to energy density, is an important task in a battery. However, it is difficult to develop a practical prediction model because relevant data is small and important information including structure, regarded as a good modality for predicting properties of materials, is barely known except composition. Inspired by these points, we propose a pretraining method utilizing a contrastive learning between compositions and structures(CLCS), which can improve the performance of voltage prediction task using only compositions of materials. First, we pretrained an composition encoder through contrastive learning between composition and structure representations, extracted by a transformer encoder and a graph neural network respectively, enabling the composition encoder to learn information associated with structures. Then, we transferred the composition encoder to a downstream task of predicting the average voltage with compositions. The performance of transferred model exceeds one of a model without pretraining by 9.7%. Also, with attention score analysis, we discovered that the transferred composition encoder focuses on lithium more than other elements in lithium-transition metal-oxygen systems compared to the composition encoder without pretraining.

1 Introduction

As Li-ion battery technology has evolved, it has gone from being expected to be the new generation energy source to being used in a variety of applications such as electric vehicles and energy storage system.[1; 2; 3] As a result, a volume of work is being conducted on cathode materials to increase energy density, which is one of the key factors of a battery. However, it is time-consuming to conduct experiments by synthesizing materials, or to calculate it using computational chemistry. With the recent increase in computational power and advent of databases such as ZINC[4], OQMD[5] and

37th Conference on Neural Information Processing Systems (NeurIPS 2023).

Materials project[6], deep learning is being applied to chemistry, such as predicting properties of molecules[7; 8], crystals [9; 10; 11] to accelerate development of molecules and materials. However, the application of deep learning in the battery field is not active due to the lack of relevant data, and important information including structure is barely known except composition information[12]. In general, the structures contain more information of materials than the compositions, which is useful to predict the properties and Bartel et al.[13] demonstrated that structure-based models outperform composition-based models through a comprehensive comparison. The same phenomenon is found in previous studies[14; 15; 16; 17] which applied deep learning to predict the voltage of cathode materials in the Materials Project. Zhang et al.[16] and Louis et al. [17] designed prediction models employing a crystal graph convolutional neural network(CGCNN)[18] and a global attention-based graph convolutional neural network(GATGNN)[19]. Since both graph-based models reflect topologies of materials well, they perform better than a model[15] using only compositions. However, structure based models are often impractical for experimental data from real-world battery experimenters, as they frequently lack essential structure information. The acquisition of crystal structure data requires advanced analytical techniques such as X-ray diffraction(XRD), neutron diffraction, and transmission electron microscopy. These methods necessitate the use of expensive equipment and specialized expertise, which may not be available in all research facilities. Moreover, accurately determining structures is a time-consuming and resource-intensive process.
In this work, in order to address two issues of little voltage data and missing structure information, we propose a pretraining method utilizing contrastive learning between compositions and structures (CLCS), enabling the composition encoder to learn information associated with structures. When the pretrained composition encoder was transferred to a task to predict the voltage from composition, the prediction performance was improved than the one without pretraining. With attention score analysis, we discovered that the transferred composition encoder focuses on lithium more than other elements in lithium-transition metal-oxygen systems compared to the one without the pretraining.

2 Methods

2.1 CLCS Pretraining

We devised a pretraining method utilizing contrastive learning between two modalities, compositions and structures,(CLCS) which enables compositional encoder to learn information associated with structures. Overview of CLCS can be seen in Figure1(a). We trained a composition encoder and a structure encoder jointly to predict correct pairings of (structure, composition) in batches by utilizing InfoNCE loss[20], similar to Contrastive language-image pretraining(CLIP)[21]. If representations of structures are similar, representations of compositions will be embedded similarly. Conversely, if they differ, representations of compositions will be embedded at a distance. Compositionally restricted attention-based network(CrabNet)[11] and directional message passing network(DimeNet)[8] are employed to get the representations of compositions and structures, respectively. When we implemented the CrabNet, cls token is added to the existing CrabNet model and considered as a representation of a composition. Pretraining was performed on two datasets, one with lithium oxides and the other with all materials in the Materials Project. The model, when pretrained on lithium oxides, could potentially acquire specific knowledge essential for lithium-ion batteries, owing to its specialized pretraining dataset. This targeted approach is especially pertinent considering that a substantial portion of Li-ion batteries, 2,260 out of every 2,440 pairs, contain oxides. Such specificity could be beneficial for downstream tasks, such as predicting the average voltage of Li-ion batteries. Conversely, a model pretrained on a diverse array of materials could develop a more expansive understanding of material properties. By undergoing pretraining with all materials, the model is equipped to identify the relationships between structures and compositions within a wide range of crystals, extending beyond the scope of just battery materials. Details of a pretraining method are described in section 5.3.

2.2 Transfer the pretrained composition encoder to a voltage prediction task

Transfer learning [22] is a technique used in many fields to improve performance by retraining a model on a specific task with generalized information learned from a large dataset. This showed the potential to overcome the limited number of material data. [23; 24] Therefore, we transferred the pretrained composition encoder to a downstream task of predicting average voltage of 2,286 batteries in the Materials project. According to the method of calculating the average voltage[25], described in

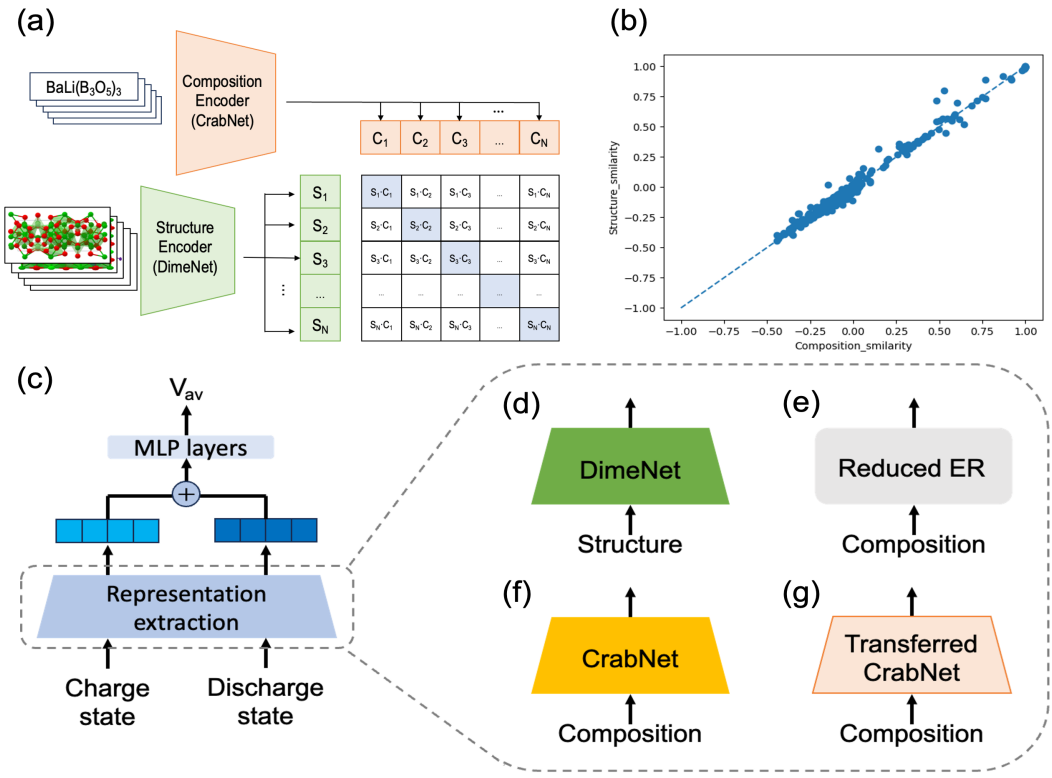

Figure 1: The overview of (a) CLCS pretraining. (b) A scatter plot was created to show the compositional and structural similarities of materials in the lithium oxides pretraining dataset, compared to a reference material(mp-757135), using the test set.(We used cosine similarity to measure similarity.) The overview of (c) the average voltage prediction. (d), (e), (f), and (g) represent S2P, Reduced ER, C2P and Transferred C2P respectively.(More details can be found in Section 3.2.)

section5.1, average voltage depends on the states of the charge and discharge. The compositions of both states were embedded through the pretrained encoder, and these were concatenated and passed through multi-layer perceptron(MLP) to predict the voltage.

3 Experimental setup and Results

3.1 CLCS pretraining

To verify CLCS was trained as intended, a scatter plot was created to show the compositional and structural similarities of materials in the lithium oxides pretraining dataset, compared to a reference material(mp-757135), using the test set.(Figure1(b)). It showed that composition representations of materials which are structurally similar are similar, and vice versa. This confirms that the composition encoder and the structure encoder learned the association between compositions and structures.

3.2 Evaluation of voltage prediction models

To demonstrate our CLCS pretraining method, we compared prediction performances of four representation extraction methods shown in Figure1(c)-(g):structure encoder in DimeNet(S2P), engineered composition representations using matminer and reduced by PCA (Reduced ER), composition encoder in CrabNet(C2P) and composition encoder transferred from CLCS pretraining(transferred C2P). In the case of transferred C2P, two models pretrained with lithium oxides and all materials are transferred to the downstream task. And for the engineered representation, we tried to reproduce the model of the previous study[15]. The details of implementation can be seen in the 5.3. In Table 1, the prediction performances of all models are recorded. The models can be divided into two parts

Table 1: The evaluation of the five models for predicting the average voltage is recorded in terms of mean absolute error with the test set. Error bars correspond to the standard deviation of the five trained models with different random seeds. "-" in the "Pretraining data" column indicates that the model was trained using random initialization instead of transferring from the pretraining step.

Models	Input type	Pretraining data	MAE	R2 Score
S2P	Structure	-	0.323 ± 0.015	0.740 ± 0.028
Reduced ER [15]	Composition	-	0.505 ± 0.015	0.502 ± 0.030
C2P	Composition	-	0.402 ± 0.008	0.615 ± 0.024
Transferred C2P	Composition	Li Oxides	$\mathbf{0.371 \pm 0.012}$	$\mathbf{0.670 \pm 0.023}$
Transferred C2P	Composition	All materials	$\mathbf{0.363 \pm 0.018}$	$\mathbf{0.704 \pm 0.032}$

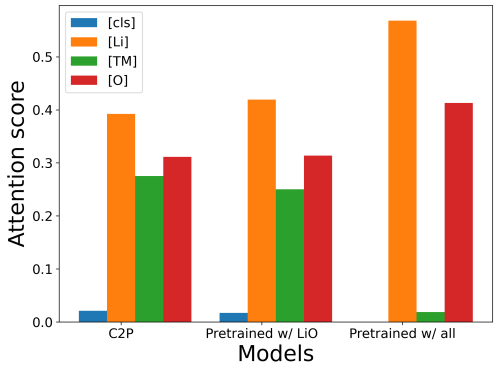

Figure 2: Given the Li-TM-O ternary system as input, the scores referring to each token were plotted as bars to create context vectors from the transformer encoders of the three models (C2P, transferred C2P pretrained with lithium oxides and transferred C2P pretrained with all materials).

according to the type of input, and S2P showed the lowest mean absolute error(MAE), as shown in the previous study[17]. However, as mentioned earlier, composition-based prediction models can be used practically in various cases, the performance of structure-based model is used as a reference which shows structure information is important. When focusing on cases where the input type is a composition, prediction performance of C2P using CrabNet is better than the previous study[15], Reduced ER. CLCS pretraining enhances the performance of the predictive model. Especially, MAE of transferred C2P pretrained with all materials, was decreased by 9.7% compared to C2P. Therefore, it has been numerically demonstrated that CLCS is a viable method when data is small and structural information is not available. Some example results of voltage prediction using the trained models(C2P and Transferred C2P) are illustrated in Table 4.

3.3 Attention score analysis

To investigate the impact of CLCS pretraining, we examined the attention scores of the transformer encoders from C2P, transferred C2P pretrained with lithium oxides and all materials and denoted them as C2P, pretrained with LiO and pretrained with all, respectively. The analysis is conducted with lithium-transition metal(TM)-oxygen ternary systems, commonly used in practice, and the used TMs are Ti, V, Cr, Mn, Fe, Co, Ni, and Cu. Attention scores, where the cls token focuses on itself, Li, TM, and O tokens, indicate what the models are focusing on and are plotted in Figure2. The methodology for extracting the attention scores is described in 5.5. In the case of C2P, attention scores on Li, TM, and O tokens are similar, whereas in the case of transferred C2P, the scores of Li and O tokens increased and the scores of others decreased. Higher attention scores for the Li token would lead to better predictions of the average voltage. Because the average voltage is closely related to the energy difference resulting from the intercalation and deintercalation of lithium within the structure.

4 Conclusion

We proposed utilizing the composition information that includes the association with the structures from contrastive learning between compositions and structures(CLCS) by transfer learning to the other task. In the voltage prediction task with compositions, the MAE of the transferred model decreased by 9.7% than the one without pretraining. In the Li-TM-O systems, Li token in the composition encoder have higher attention scores than the others, as the structural information is incorporated in the CLCS pretraining. To our knowledge, this is the first approach to train the composition encoder to learn information associated with structures, which improves the prediction performance in the task where compositions are only available. While a downstream task was limited to the voltage prediction of batteries in this study, it would be interesting to see if this methodology holds up in other tasks; if it does, it is more practical than structure-based prediction models for researchers conducting experiments in the real world. During the pretraining, duplicates with the same composition were removed by using only the most stable structures based on E_{hull}. To improve the model, it would be beneficial to either use structures similar to cathode materials or develop a method for learning a distribution that reflects the variability of structures within the same composition. Furthermore, there are more ways to incorporate structure information. For example, instead of training the structure encoder and composition encoder together in the pretraining, we could distill the knowledge from the pretrained structure encoder to the composition encoder. However, we leave this as a future work because it takes time to test which labels to pretrain the structure encoder with. In this work, we focused on the fact that we improved the performance with simple contrastive learning despite not carrying structure information from the pretrained structure encoder.

5 Supplementary Material

5.1 Calculation method of average voltage

The average voltage of intercalation electrode is derived from the previous study[25].

$$V_{av} = -\frac{\Delta G}{\Delta x F} \tag{1}$$

ΔG, Δx, and F denote the change of the Gibbs free energy, number of Li-ions between the charge state and the discharge state, and Faraday constant. Since V_{av} depends on the charge state and the discharge state, both states are used as inputs to increase prediction performance.

5.2 Dataset preparation

Two datasets were used in this study: the first is a dataset for the pretraining CLCS, and the second is a dataset for average voltage prediction. The Python Materials Genomics(pymatgen) library[26], open-source and widely used library, was used to download data from the Materials project and to curate the data. The data preparation process is as follows:

Pretraining dataset Dataset including lithium oxides only and including all materials from the Material Project database are used for pretraining. Each dataset was cleaned by removing unstable structures in each compositions based on the energy above hull, E_{hull}. The number of lithium oxides and all materials used for pretraining were 6,530 and 105,583, and the train set, validation set, and test set were divided in the ratio of 8:1:1.

Average voltage prediction dataset 2,440 Li-ion battery data was downloaded from the materials Project. In most cases, battery data consists of a single step, but some consist of multiple steps. In all cases, we considered the states with the least and the most lithium content as the charge and discharge states, respectively, and used the corresponding average voltage values. We used interquartile range(IQR) method to remove outliers. The IQR represents the range between the first quartile(Q1) and the third quartile(Q3) of a dataset and values that do not fall between (Q1 - 1.5 * IQR) and (Q3 + 1.5 * IQR) are considered outliers and are removed. The remaining dataset consisted of 2,286 data points. These were then divided into train, validation, test sets at an 8:1:1 ratio.

5.3 Implementation of average voltage prediction models

In the experimental setup and results, there are CLCS pretraining architecture and four average voltage prediction models(S2P, C2P, Transferred C2P and Reduced ER). All models are trained with the same machine, which has an AMD EPYC 7413@2.65 GHz, 1TB of RAM, and NVIDIA A40 GPU.

CLCS pretraining A contrastive learning between compositions and structures was performed to the make the composition encoder learn information associated with structures. In this process, CrabNet and DimeNet were used to encode the composition information and the structure information, respectively. The hyperparameters of the models are described in Table 2 and InfoNCE[20] was used for training.

Average voltage prediction models All predictive models used both charge and discharge states as inputs. The difference between models is how to encode the inputs. The prediction model using structure information encoded the charge and discharge states with the DimeNet[8]. DimeNet is a model based on the graph neural network which uses a directional information by transforming messages based on the angle between them. By concatenating representations of two atoms and distance between them, a message embedding is generated, which then undergoes updates and aggregations through several interaction blocks. Afterwards, messages directed towards each atom are combined to produce an atom-wise output. An embedding of structure can be represented by summing up the atom embeddings obtained from each interaction layer. By leveraging the distances and angles which are invariant to translation and rotation, this model can encode structures in a principled and effective manner. For C2P and Transferred C2P, the composition information was encoded with the CrabNet[11]. CrabNet is a model based on transformer encoders[27], which tokenize each element of a composition and encode fractions with sinusoidal functions. Embeddings of each element are combined with their fractional encodings and then processed using transformer encoders. This approach demonstrated superior performance in predicting properties of materials based on their compositions, which indicates compositions are effectively encoded by this model. For

Table 2: Hyperparameters of CLCS pretraining

model	hyperparameters	value
CrabNet	the number of transformer encoder layer	3
	the number of heads in the multiheadattention models	4
	the number of expected features in the input	64
	dimension of feedforward	2048
	dropout	0.1
	dimension of output embedding	64
DimeNet	the number of blocks	4
	the number of spherical	7
	the number of spherical	6
	the number of radial	6
	cutoff radius	7
	the max number of neighbors	20
	envelope of exponents	5
	dimension of output embedding	64
common	batch size	256
	learning rate	1e-4

Table 3: Hyperparameters of training the voltage prediction models

model	hyperparameters	value
CrabNet	the number of transformer encoder layer	3
	the number of heads in the multiheadattention models	4
	the number of expected features in the input	64
	dimension of feedforward	2048
	dropout	0.1
	dimension of output embedding	64
DimeNet	the number of blocks	4
	the number of spherical	7
	the number of spherical	6
	the number of radial	6
	cutoff radius	7
	the max number of neighbors	20
	envelope of exponents	5
	dimension of output embedding	64
common	batch size	64
	learning rate	1e-4

all models, the representations of the charge and discharge states were concatenated and fed into MLP layers for predictions. Reduced ER model was reproduced from the previous study.[15] Since the code was not open to the public, so we tried to reproduce it as written in the paper. Matminer[28] was used to create a composition representation. Feature normalization and dimension reduction with PCA were conducted using scikit learn library[29]. However, this method has limitation that if composition includes uranium(U), representaion can not be made with Matminer. Therefore, we remove data("mp-759350_Li") including uranium when we train Reduced ER. The overall hyperparameters of the models are described in Table 3 and mean absolute error(MAE) was used for training.

5.4 Example results of voltage prediction

Example results of voltage prediction using the trained models, namely C2P and Transferred C2P pretrained with all materials, are illustrated to showcases the effectiveness of CLCS pretraining. These examples demonstrate the impact of comprehensive pretraining. Examples are randomly selected in the test set.

Table 4: Example results of voltage prediction using the trained models, namely C2P and Transferred C2P pretrained with all materials, are illustrated to showcases the effectiveness of CLCS pretraining. These examples demonstrate the impact of comprehensive pretraining.

Battery ID	Prediction of C2P(V)	Prediction of Transferred C2P(V)	Ground truth(V)
mp-18001_Li	4.886	3.949	1.504
mp-26095_Li	5.075	5.043	4.913
mp-26995_Li	4.862	4.806	4.803
mp-35171_Li	1.891	1.765	1.694
mp-755702_Li	1.970	2.004	2.070
mp-758577_Li	5.038	5.031	5.052

5.5 Extracting attention scores

We investigated effects of each elements in the composition to overall context(cls token) by extracting attention scores in the first transformer encoder layer in the Figure 2. Because each token in the first multi-headed attention layer of a CrabNet has relatively pure elemental information before interacting with each other by its own self attention, the interactions of how much they affect each other can be observed clearly. Inspired by promising cathode materials such as lithium manganese oxide(LMO) and lithium cobalt oxide(LCO), the attention score analysis was conducted with materials consisting of lithium and transition metal and oxygen(Li-TM-O) systems. Transition metals from the periodic table in the same row as Mn and Co were used. The list includes: Ti, V, Cr, Mn, Fe, Co, Ni and Cu. To examine how much the cls token representing the materials is influenced by other tokens, we compared the average attention scores obtained across 189 Li-TM-O systems out of 2,286 data points in each model. The counts of systems for each transition metal are 19, 29, 17, 35, 32, 23, 24 and 10, respectively. Although Li-Sc-O and Li-Zn-O systems are in the same row as Mn and Co on the periodic table, they do not exist in the dataset and therefore were not used.

References

[1] Michel Armand and J-M Tarascon. Building better batteries. *nature*, 451(7179):652–657, 2008.

[2] Naoki Nitta, Feixiang Wu, Jung Tae Lee, and Gleb Yushin. Li-ion battery materials: present and future. *Materials today*, 18(5):252–264, 2015.

[3] Bruno Scrosati, Jusef Hassoun, and Yang-Kook Sun. Lithium-ion batteries. a look into the future. *Energy & Environmental Science*, 4(9):3287–3295, 2011.

[4] John J Irwin, Teague Sterling, Michael M Mysinger, Erin S Bolstad, and Ryan G Coleman. Zinc: a free tool to discover chemistry for biology. *Journal of chemical information and modeling*, 52(7):1757–1768, 2012.

[5] Scott Kirklin, James E Saal, Bryce Meredig, Alex Thompson, Jeff W Doak, Muratahan Aykol, Stephan Rühl, and Chris Wolverton. The open quantum materials database (oqmd): assessing the accuracy of dft formation energies. *npj Computational Materials*, 1(1):1–15, 2015.

[6] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL materials*, 1(1), 2013.

[7] Zhenqin Wu, Bharath Ramsundar, Evan N Feinberg, Joseph Gomes, Caleb Geniesse, Aneesh S Pappu, Karl Leswing, and Vijay Pande. Moleculenet: a benchmark for molecular machine learning. *Chemical science*, 9(2):513–530, 2018.

[8] Johannes Gasteiger, Janek Groß, and Stephan Günnemann. Directional message passing for molecular graphs. *arXiv preprint arXiv:2003.03123*, 2020.

[9] Juhwan Noh, Seoin Back, Jaehoon Kim, and Yousung Jung. Active learning with non-ab initio input features toward efficient co 2 reduction catalysts. *Chemical science*, 9(23):5152–5159, 2018.

[10] Qiuling Tao, Pengcheng Xu, Minjie Li, and Wencong Lu. Machine learning for perovskite materials design and discovery. *npj Computational Materials*, 7(1):23, 2021.

[11] Anthony Yu-Tung Wang, Steven K Kauwe, Ryan J Murdock, and Taylor D Sparks. Composition-ally restricted attention-based network for materials property predictions. *Npj Computational Materials*, 7(1):77, 2021.

[12] Guanyu Wang, Tom Fearn, Tengyao Wang, and Kwang-Leong Choy. Machine-learning approach for predicting the discharging capacities of doped lithium nickel–cobalt–manganese cathode materials in li-ion batteries. *ACS central science*, 7(9):1551–1560, 2021.

[13] Christopher J Bartel, Amalie Trewartha, Qi Wang, Alexander Dunn, Anubhav Jain, and Gerbrand Ceder. A critical examination of compound stability predictions from machine-learned formation energies. *npj computational materials*, 6(1):97, 2020.

[14] Rajendra P Joshi, Jesse Eickholt, Liling Li, Marco Fornari, Veronica Barone, and Juan E Peralta. Machine learning the voltage of electrode materials in metal-ion batteries. *ACS applied materials & interfaces*, 11(20):18494–18503, 2019.

[15] Isaiah A Moses, Rajendra P Joshi, Burak Ozdemir, Neeraj Kumar, Jesse Eickholt, and Veronica Barone. Machine learning screening of metal-ion battery electrode materials. *ACS Applied Materials & Interfaces*, 13(45):53355–53362, 2021.

[16] Xiuying Zhang, Jun Zhou, Jing Lu, and Lei Shen. Interpretable learning of voltage for electrode design of multivalent metal-ion batteries. *npj Computational Materials*, 8(1):175, 2022.

[17] Steph-Yves Louis, Edirisuriya M Dilanga Siriwardane, Rajendra P Joshi, Sadman Sadeed Omee, Neeraj Kumar, and Jianjun Hu. Accurate prediction of voltage of battery electrode materials using attention-based graph neural networks. *ACS Applied Materials & Interfaces*, 14(23):26587–26594, 2022.

[18] Tian Xie and Jeffrey C Grossman. Crystal graph convolutional neural networks for an accurate and interpretable prediction of material properties. *Physical review letters*, 120(14):145301, 2018.

[19] Steph-Yves Louis, Yong Zhao, Alireza Nasiri, Xiran Wang, Yuqi Song, Fei Liu, and Jianjun Hu. Graph convolutional neural networks with global attention for improved materials property prediction. *Physical Chemistry Chemical Physics*, 22(32):18141–18148, 2020.

[20] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *arXiv preprint arXiv:1807.03748*, 2018.

[21] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PMLR, 2021.

[22] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2009.

[23] Hironao Yamada, Chang Liu, Stephen Wu, Yukinori Koyama, Shenghong Ju, Junichiro Shiomi, Junko Morikawa, and Ryo Yoshida. Predicting materials properties with little data using shotgun transfer learning. *ACS central science*, 5(10):1717–1730, 2019.

[24] Geun Ho Gu, Jidon Jang, Juhwan Noh, Aron Walsh, and Yousung Jung. Perovskite synthesizability using graph neural networks. *npj Computational Materials*, 8(1):71, 2022.

[25] MK Aydinol, AF Kohan, Gerbrand Ceder, Kang Cho, and JJPRB Joannopoulos. Ab initio study of lithium intercalation in metal oxides and metal dichalcogenides. *Physical Review B*, 56(3):1354, 1997.

[26] Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L Chevrier, Kristin A Persson, and Gerbrand Ceder. Python materials genomics (pymatgen): A robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, 2013.

[27] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[28] Logan Ward, Alexander Dunn, Alireza Faghaninia, Nils ER Zimmermann, Saurabh Bajaj, Qi Wang, Joseph Montoya, Jiming Chen, Kyle Bystrom, Maxwell Dylla, et al. Matminer: An open source toolkit for materials data mining. *Computational Materials Science*, 152:60–69, 2018.

[29] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.